

Gene Search

Introduction

[Home](#) > [Find Genes](#) > [Gene Search](#)

Gene Search

Gene Search is a way to search, browse and select genes in IMG. It can be accessed via the second-level menu of **Find Genes**, as shown in Figure 1.

img INTEGRATED MICROBIAL GENOMES

IMG Home Find Genomes **Find Genes** Find Functions Compare Genomes Analysis Cart My IMG Using IMG

Home > Find Genes

Gene Search

Find genes in selected genomes by keyword.

Keyword:

Filters: Gene Product Name (inexact)

☐ Additional Output Columns
(not applied to 'Pfam Domain', 'Protein Regular Expression Pattern' searches)

Genome Filter

Domains: (B)acteria, (A)rchaea, (E)ukarya, (P)lasmiids, (V)iruses.
Genome Completion: [F]inished, [P]ermanent Draft, [D]raft.

Seq. Status Domain

All Finished, Permanent Draft and Draft All

☒ List ☐ Tree

Acidilobus saccharovorans 345-15 (A)[F]
Aciduliprofundum boonei T469 (A)[D]
Aciduliprofundum boonei T469 (A)[F]
Aeropyrum pernix K1 (A)[F]
Archaeoglobus fulgidus DSM 4304 (A)[F]
Archaeoglobus profundus DSM 5631 (A)[F]
Caldivirga maquilingensis IC-167 (A)[F]
Candidatus Korarchaeum cryptofilum OPF8 (A)[F]
Candidatus Methanoregula boonei 6A8 (A)[F]
Candidatus Methanoregula boonei 6A8 (A)[F]

Examples

- "kinase" as a Gene Product Name gets "Shikimate kinase".
- "fusA" as a Gene Symbol (exact) gets only genes with the symbol "fusA".
- "b0179" as a Locus Tag gets only genes with the locus tag "b0179".
- "NP_414721" will retrieve genes with that GenBank accession.
- "638154502" for IMG Gene Object Identifier will directly retrieve gene with that identifier. (Genome selection is ignored.)
- "16128172" for NCBI GI Number will retrieve the corresponding IMG gene. (Genome selection is ignored.)
- (Try the "kinase" example or See IMG Term browser for valid IMG terms.)
- (Generally, ID searches are 'exact' searches. Search on descriptive fields are 'inexact', i.e. substring searches. Except when indicated otherwise, all searches are exact here.)
- "Acyl_transf_1, !PP-binding" as a comma separated list specification will find genes that have the *Acyl_transf_1* domain and not *PP-binding* in Pfam domain search.
- "pfam00698, pfam00109, !pfam00668" as a comma separated list specification will find genes that have the pfam domain hits *pfam00698*, *pfam00109*, and not *pfam00668* in Pfam domain search.

Figure 1: Gene Search.

Keyword Search

User can search for genes containing a keyword in a specific field of the IMG database. A total of 12 search filters are available for selection, as shown in Figure 2.

The default filter is “Gene Product Name”. User can also select filters such as “Gene Symbol”, “Locus Tag”, “GenBank Accession”, “NCBI GI Number”, “IMG Gene Object ID”, “IMG Term and Synonyms”, “SEED Product Name/Subsystem”, “Is Obsolete Gene” and “Is Pseudo Gene”. Further, user can find genes by entering Pfam ID(s) against Pfam Domain or entering a regular expression match against proteins.

Keyword:

Filters: Gene Product Name (inexact) ▼

☐ Add (not applied to searches)

- Gene Product Name (inexact)
- Gene Symbol (list)
- Locus Tag (list)
- GenBank Accession (list)
- NCBI GI Number (list)
- IMG Gene Object ID (list)
- IMG Term and Synonyms (inexact)
- SEED Product Name/Subsystem (inexact)
- Is Obsolete Gene ("Yes" or "No")
- Is Pseudo Gene ("Yes" or "No")
- Pfam Domain Search (list)
- Protein Regular Expression Pattern (inexact)

Figure 2: Gene Search Filters.

In addition to search filters, user is also provided the capability to select and display “Additional Output Columns” in result page, as shown in Figure 3:

Keyword:

Filters: Gene Product Name (inexact) ▼

☐ Additional Output Columns

- ☐ Gene Symbol
- ☐ GenBank Accession
- ☐ Chromosome
- ☐ Start Coord
- ☐ End Coord
- ☐ Strand
- ☐ DNA Sequence Length
- ☐ AA Sequence Length
- ☐ Locus Type
- ☐ Is Pseudogene
- ☐ Is Obsolete
- ☐ Add Date
- ☐ Scaffold Object ID

(not applied to 'Pfam Domain', 'Protein Regular Expression Pattern' searches)

Figure 3: Additional Output Columns for Selection and Display.

These additional display columns are arranged into the following five groups:

1. Gene Symbol; GenBank Accession
2. Chromosome; Start Coord; End Coord; Strand; DNA Sequence Length; AA Sequence Length; Locus Type
3. Is Pseudogene; Is Obsolete
4. Add Date
5. Scaffold

Genome Filter

The Genome Filter is provided to limit the scope of search. The search, except “NCBI GI Number” and “IMG Gene Object ID” filters, is conducted on the genomes that user has saved through the **Genome Browser**. If user saves nothing, by default, it's all genomes in IMG database. However, user can override the selection via the **Genome Filter** provided in the page.

Since genome selection is ignored in “NCBI GI Number” and “IMG Gene Object ID”. The **Genome Filter** section turns hidden when user selects these two filters, and will display again when user selects other filters.

For more information about **Genome Filter**, go to “[Home > User Guide > Site Map](#)”.

Search Specifications

All searches are case-insensitive. In general, ID or number or symbol searches are 'exact' searches. Searches on descriptive fields are 'inexact', that is, the keyword is used as a word or part of a word.

For example, (i) searching for "fusA" with the filter set to "Gene Symbol" only gets genes with the symbol "fusA"; searching for "bsu02690" with the filter set to "Locus Tag" only gets genes with locus tag "BSU02690"; (ii) searching for "kin" with the filter set to "Product Name" gets "Shikimate kinase," etc.

User can use a percent sign (%) as a wildcard in the middle of a keyword. The results will include any genes with zero or more additional characters at that position. For example, "hydro%ase" will get results with "hydrolase" and "hydrogenase". If user wants only a single character of the keyword to be variable, type an underscore (_) in that position. Searching for "hydro_ase" will get results with "hydrolase," not "hydrogenase."

For regular-expression protein searches, standard regular expression constructs are used against a sequence of amino acid residues. Common constructs include:

- . Matches and single amino acid
- \$ Matches end of sequence.
- ^ Matches beginning of sequence.

- Matches zero or more occurrences of the preceding residue.
- + Matches one or more occurrences of the preceding residue.
- [] Matches any of residues between brackets.
- [^] Matches any of residues not between brackets.